Neurocomputing 137 (2014) 157–164

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

A comparative study and improvement of two ICA using reference signal methods

Jian-Xun Mi [a,b,*], Yong Xu [c,d]

[a] College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing, 400065, China
[b] Chongqing Key Laboratory of Computational Intelligence, Chongqing University of Posts and Telecommunications, Chongqing, 400065, China
[c] Bio-Computing Research Center, Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen
[d] Key Laboratory of Network Oriented Intelligent Computation, Shenzhen, Guangdong Province, China

ARTICLE INFO

Article history:
Received 17 December 2012
Received in revised form
6 March 2013
Accepted 23 March 2013
Available online 16 February 2014

Keywords:
Independent component analysis
ICA with reference signal
Wiener filter

ABSTRACT

In this paper, we present a comparative assessment of two ICA using reference signal methods. Independent Component Analysis (ICA) using reference signal is a useful tool for extracting a desired independent component (IC). Reference signal is served as *a priori* information to conduct the one-unit ICA to converge to the local extreme point related to a desired IC. There are two methods can perform ICA using reference signal, namely ICA with reference (ICA-R) and fast ICA with reference signal (FICAR). This paper intends to fill in a gap in the previous studies to give comparisons of those methods systematically. Moreover, we also provide useful improvements to the two methods. Firstly, we propose a improved algorithm to fix the flaw in the previous FICAR. Secondly, for ICA-R, we propose a criterion to select appropriate distance measurement.

1. Introduction

To discover interesting components from observed data is an unfailing hotpot in the area of computer application [1–6]. One of the most widely used methods is principal component analysis (PCA) [7,8] and its varieties, such as Kernel PCA [9–12], 2D-PCA [13,14], matrix-based complex PCA (MCPCA) [15]. PCA finds some principal components, which minimizes the squared error between the original data and its reconstruction, by using an orthogonal transformation. ICA, another famous components mining method, tries to reveal independent sources which underlie the observed multichannel data. For the independence of the recovered components, ICA becomes an attractive technique for application in the field of blind source separation. In some practical applications, only a subset of ICs is of interest. A simple and naive way is to using classical ICA method to compute all ICs first and then use available *a priori* information, namely some properties and/or requests on the results, to select some required ICs from them. Nevertheless, in this scheme, undesired ICs are computed at first and discarded finally, which is very consumptive. Especially, in many cases, observed data consist of a very large number of underlying components, which makes exhaustive computation of all ICs unbearable or unfeasible.

Before ICA method, the Wiener filtering (WF), a second-order approach, has been widely used to extract a source from observed data. Usually, *a priori* information is required to produce a reference signal. WF extracts the underlying soure which is the closest one to the reference signal based on second-order distance criteria. However, this technique is hardly to recover independent sources, in higher-order statistical sense. The reason is that if more than one IC has correlations with the reference signal, which is usually true, the output by WF cannot be statistically independent source. Thus, WF is hard to extract ICs.

A framework called constrained ICA (cICA) was proposed by Lu and Rajapakse in [16]. The cICA is a technique to extract IC by incorporating extra requirements into ICA algorithm using equality and inequality constraints. The extension work, which is called ICA with reference (ICA-R) [17], was to incorporate the reference signal into the ICA contrast function under cICA framework. So far, ICA-R has been used in many applications, such as electroencephalograms (EEG) data processing [18], functional magnetic resonance imaging (fMRI) analysis [19,20], computer vision [21], etc. [22]. And many papers improved the performance of the original ICA-R algorithm [23–28].

FICAR was first developed in biomedical engineering society [29] which was used to cancel cardiac artifacts from a magnetoencephalogram (MEG). FICAR is a two-stage method which first uses the reference signal to form an initial weight vector by

* Corresponding author at: College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing, 400065, China. Tel.: +86 13527500800.
E-mail address: mijianxun@gmail.com (J.-X. Mi).

http://dx.doi.org/10.1016/j.neucom.2013.03.070
0925-2312

Wiener filter and then extracts an IC by one-unit Fast ICA with the Wiener weight. FICAR was also used in EEG analysis [30], Brain Computer Interface (BCI) techniques [31], etc.

Compared with classical ICA methods, both ICA-R and FICAR take advantage of *a priori* information to recover the desired IC so that they avoid computing uninteresting ICs. For ICA-R, *a priori* information is added into ICA-R as an inequality constraint to form an augmented Lagrangian function, and the reference signal works during searching the function's global minimum. FICAR uses the reference signal only in its first stage to construct the Wiener weight as the initial search vector for the following processing. According to [29], if the Wiener weight vector is close enough to the solution vector associated with the interesting IC, one-unit FastICA can produce the correct IC at the second stage. That is to say, FICAR have already exploited the reference signal before solving the ICA optimization problem. Although the reference signal is used differently, both methods are capable of extract the desired IC without exhaustively calculation of all ICs. However, there is no study to compare ICA-R and FICAR theoretically and experimentally. In this paper, we present a comparative assessment of the mentioned two ICA using the reference signal techniques.

The major contributions of the paper are as follows. We present and analyze the differences between ICA-R and FICAR. The factors which affect the convergence speed of both methods were given. Then, we improve the algorithm for FICAR and propose the approach of selecting an appropriate distance measurement for ICA-R to reasonably approximate the Hessian matrix.

The manuscript is organized as follows. Section 2 introduces Wiener filtering. Section 3 first defines the ICA using reference signal problem, then describes FICAR algorithm and ICA-R algorithm, and finally presents a summary of differences between them. We propose a new method for FICAR to measure distance between the weight vector and its corresponding initiation in Section 4, as well as the principle for ICA-R to select an appropriate distance measurement. We compare two methods experimentally in Section 5. Finally, Section 6 provides conclusions.

2. Wiener filtering and ICA using reference signal problem

Before introducing the ICA using reference signal problem, we briefly review Wiener filtering technique which is a classical second-order method to extract source from multichannel observed data.

For studying source extraction problem, we assume the observed multichannel data $\mathbf{x}=(x_1,x_2,\cdots,x_n)^T$ is an instantaneous linear mixture of ICs $\mathbf{s}=(s_1,s_2,\cdots,s_m)^T$ by a matrix $\mathbf{A}$of size $n\times m$, which is

$$\mathbf{x}=\mathbf{A}\mathbf{s} \tag{1}$$

For simplicity, we address the problem of complete ICA, so that $n=m$ is assumed in the paper.

Without loss of generality, we whiten $\mathbf{x}$ to have

$$\mathbf{z}=\mathbf{V}\mathbf{x} \tag{2}$$

where $\mathbf{V}$ is an $n\times n$ whitening matrix, so that $E\{\mathbf{z}\mathbf{z}^T\}=\mathbf{I}$. In practice, the pre-whitening operation can simplify and speed up ICA algorithms. Then we let $\mathbf{B}=\mathbf{V}\mathbf{A}$, and rewrite (2) as

$$\mathbf{z}=\mathbf{B}\mathbf{s} \tag{3}$$

which is treated as the observed input signals. It is easy to know that $\mathbf{B}^T\mathbf{B}=\mathbf{I}$.

To extract a source from the mixture data, we project $\mathbf{x}$ onto a weight vector $\mathbf{w}=(w_1,w_2,\cdots,w_n)^T$ to produce the output $y=\mathbf{w}^T\mathbf{z}$. Let r be the reference signal constructed based on the available *a priori* information. Without loss of generality, we assume that all variables, s_i $(i=1,2,\cdots,n)$, r, and x_i $(i=1,2,\cdots,n)$, are with zero means and unit variances. The aim of second-order source extraction method is to find the optimum weight $\mathbf{w}^*$ to make the output $y^*=\mathbf{w}^*\mathbf{z}$ minimize the mean square error (MSE) given by $E\{(r-y^*)^2\}$. By straightforward algebra calculation [29], the optimum weight is given by

$$\mathbf{w}^*=\frac{E\{\mathbf{z}r\}}{\|E\{\mathbf{z}r\}\|} \tag{4}$$

where $\|\cdot\|$ is the Frobenious norm which constrains $\mathbf{w}^*$ to have a unit length. Then, let us focus on the output

$$y^*=\frac{E\{\mathbf{z}r\}^T\mathbf{z}}{\|E\{\mathbf{z}r\}\|}=\frac{E\{\mathbf{B}\mathbf{s}r\}^T\mathbf{B}\mathbf{s}}{\|E\{\mathbf{B}\mathbf{s}r\}\|}=\frac{E\{\mathbf{s}^Tr\}\mathbf{s}}{\|E\{\mathbf{s}r\}\|}=\mathbf{q}^T\mathbf{s}$$
$$\mathbf{q}=E\{\mathbf{s}^Tr\}/\|E\{\mathbf{s}r\}\|. \tag{5}$$

If $\mathbf{q}=\pm\mathbf{e}_i$, where $\mathbf{e}_i$ is a canonical base vector that the ith component is 1 and others are 0, the optimum output y^* is the ith independent source. However, for the reference signal is just a rough template carrying the *a priori* information of the desired IC, it is very rigorous to demand that only the desired IC has non-zero correlation with the reference signal. It is concluded in [17] that the second-order statistics is insufficient to recover independent sources from the observed mixture of ICs.

3. Methods of ICA using reference signal

In contrast to the second-order source extraction methods, ICA truly uncovers independent sources using high-order statistics. Generally speaking, ICA methods include two aspects: objective function and optimization algorithm. Several independent measurements or independent source separating criterion functions have been proposed [32,33]. Some ICA algorithms using *Kurtosis* or *Negentropy* as their objective function can extract one independent component at a time, i.e. one-unit ICA algorithm. However, note that the order of the extracted ICs is arbitrary.

3.1. Problem of ICA using reference signal

In the sense of semi-blind source extraction, available *a priori* information should be used within objective function instead of post-selecting interesting source after all ICs are recovered. Therefore, the ICA using the reference signal problem is described as: the one-unit ICA algorithm is capable of extracting an interesting IC directly under the assistance of an available reference signal. We use a neural network model to illustrate this problem in Fig. 1.

The two methods compared in this manuscript use *negentropy* as their ICA contrast function. The negentropy of a signal y is defined by

$$J(y)=H(y_{Gaus})-H(y) \tag{6}$$

where y_{Gaus} is a Gaussian random variable with the variance as y and $H(\bullet)$ denotes the entropy of a variable. As we know, for

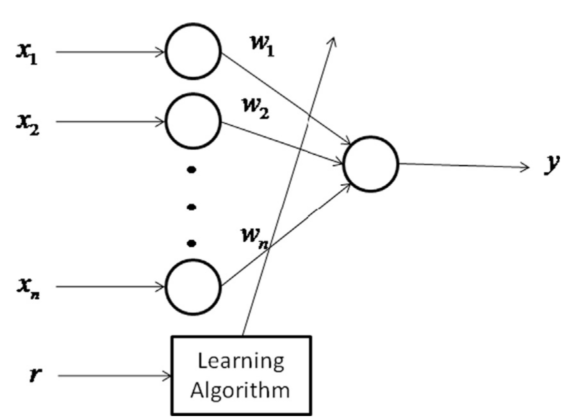

Fig. 1. Illustration of the ICA using reference signal problem by a neural network model.